## McCallum, Barry

| | |
|---|---|
| **From:** | Davis, Ben |
| **Sent:** | Wednesday, March 10, 2010 10:29 PM |
| **To:** | McCallum, Barry |
| **Subject:** | Re: archived |

We have plans for everything. You'll be impressed by what you hear.

---

**From:** McCallum, Barry
**To:** Davis, Ben
**Sent:** Wed Mar 10 21:53:26 2010
**Subject:** RE: archived
Wow! When I get back we need to discuss what's going on in the data shop. Any perpsective on how we proceed with the Plato contract next fiscal? I assume the contract has not been signed yet?

---

**From:** Davis, Ben
**Sent:** Wednesday, March 10, 2010 10:33 AM
**To:** McCallum, Barry
**Subject:** FW: archived

Barry,

We have thoroughly researched what needs to be done ▨▨▨▨▨▨▨▨▨▨▨▨ We will cover everything off without having to reach out for additional assistance.

B.

---

**From:** Power, Debbie
**Sent:** Wednesday, March 10, 2010 11:59 AM
**To:** Davis, Ben
**Subject:** FW: archived

FYI

---

**From:** ▨▨▨▨▨▨▨
**Sent:** Wednesday, March 10, 2010 11:30 AM
**To:** Power, Debbie
**Subject:** RE: archived

Thanks a lot Debbie,

We really appreciate this.

▨▨▨▨▨

-----Original Message-----
**From:** Power, Debbie
**Sent:** Wednesday, March 10, 2010 11:06 AM
**To:** Murphy, Eugene; ▨▨▨▨▨▨▨

**Cc:** Mercer, Stephen G
**Subject:** archived

Good morning ▓▓▓▓ and Eugene,

   I have archived the final weight analysis export for commercial cod ComCod2J09.wgt, ComCod3K09.wgt and ComCod3L09.wgt to dbase:[commercial]com_all_wgt.2009. The new record count is 4538.

Thanks
Debbie

*Debbie Anne Power* 🖱

**Fisheries and Oceans Canada**
NAFC, 80 East White Hills Rd.
P.O. Box 5667, St. John's, NL A1C 5X1
Telephone: 709.772.4247
E-Mail: debbie.power@dfo-mpo.gc.ca

**Peches et Océans Canada**
NAFC, 80 Rue White Hills Est.
C.P. 5667, St. John's (T-N-L) A1C 5X1
Téléphone: 709.772.4742
Courriel: debbie.power@dfo-mpo.gc.ca

## McCallum, Barry

| | |
|---|---|
| **From:** | Davis, Ben |
| **Sent:** | Wednesday, March 10, 2010 10:39 PM |
| **To:** | McCallum, Barry |
| **Subject:** | Re: archived |

Actually, we're developing a comprehensive "way forward" plan that we'll present to you. Debbie has done an outstanding job with all of the files she is managing.

B.

---

**From:** McCallum, Barry
**To:** Davis, Ben
**Sent:** Wed Mar 10 21:53:26 2010
**Subject:** RE: archived
Wow! When I get back we need to discuss what's going on in the data shop. Any perpsective on how we proceed with the Plato contract next fiscal? I assume the contract has not been signed yet?

---

**From:** Davis, Ben
**Sent:** Wednesday, March 10, 2010 10:33 AM
**To:** McCallum, Barry
**Subject:** FW: archived

Barry,

We have thoroughly researched what needs to be done ▒▒▒▒▒▒▒▒▒▒▒▒▒▒▒ We will cover everything off without having to reach out for additional assistance.

B.

---

**From:** Power, Debbie
**Sent:** Wednesday, March 10, 2010 11:59 AM
**To:** Davis, Ben
**Subject:** FW: archived

FYI

---

**From:** ▒▒▒▒▒▒▒
**Sent:** Wednesday, March 10, 2010 11:30 AM
**To:** Power, Debbie
**Subject:** RE: archived

Thanks a lot Debbie,

We really appreciate this.

▒▒▒▒▒

-----Original Message-----
**From:** Power, Debbie

1

000003

**Sent:** Wednesday, March 10, 2010 11:06 AM
**To:** Murphy, Eugene;
**Cc:** Mercer, Stephen G
**Subject:** archived

Good morning ████ and Eugene,

   I have archived the final weight analysis export for commercial cod ComCod2J09.wgt, ComCod3K09.wgt and ComCod3L09.wgt to dbase:[commercial]com_all_wgt.2009.  The new record count is 4538.

Thanks
Debbie

*Debbie Anne Power*

**Fisheries and Oceans Canada**
NAFC, 80 East White Hills Rd.
P.O. Box 5667, St. John's, NL A1C 5X1
Telephone:  709.772.4247
E-Mail: debbie.power@dfo-mpo.gc.ca

**Peches et Océans Canada**
NAFC, 80 Rue White Hills Est.
C.P. 5667, St. John's (T-N-L) A1C 5X1
Téléphone: 709.772.4742
Courriel: debbie.power@dfo-mpo.gc.ca

2

## McCallum, Barry

| | |
|---|---|
| **From:** | Davis, Ben |
| **Sent:** | Sunday, March 21, 2010 10:13 AM |
| **To:** | McCallum, Barry |
| **Subject:** | Re: Plato Contract |

Ok. I'll ask her to set an appt.

Just to clarify something that Debbie pointed out in my email below - the time requirement is 9-12 months to complete the work, not 12-21.

B.

---

**From:** McCallum, Barry
**To:** Davis, Ben
**Sent:** Sun Mar 21 09:30:02 2010
**Subject:** RE: Plato Contract
Ok. Set something up with Marlene for early in the week.

Barry

---

**From:** Davis, Ben
**Sent:** Sunday, March 21, 2010 8:18 AM
**To:** McCallum, Barry
**Subject:** Re: Plato Contract

Sure. No problem.

B.

---

**From:** McCallum, Barry
**To:** Davis, Ben
**Sent:** Sun Mar 21 09:16:47 2010
**Subject:** RE: Plato Contract
Ben,

I would like to meet with you and Debbie asap to discuss further. In the meantime, I have asked Krystine to review the contract so that we fully understand our options.

Barry

---

**From:** Davis, Ben
**Sent:** Friday, March 19, 2010 8:01 AM
**To:** McCallum, Barry
**Cc:** French, Krystine; Power, Debbie
**Subject:** RE: Plato Contract

Barry,

1

We can discuss this further as you wish. Now that Debbie is the project manager for the Plato Contract, we have exponentially increased the pace of achieving milestones and deliverables. Originally, there was an expectation of having to run this contract for another 2-3 years but since Debbie has taken this over, we are down to a hard requirement of one year and possibly nine months. She has been engaging the scientists and biologists on the real requirements for the database and cooperation is excellent and enthusiasm has grown because people can see an endpoint with a product that will be meaningful to the Branch.

The bottom line is we need the Plato programmers for another year. Concurrently, we have developed a business case for the structure and HR needs to support our information and data requirements into the future. We would like to present that business case when you have time to hear it.

Ben

---

**From:** French, Krystine
**Sent:** Thursday, March 18, 2010 7:16 PM
**To:** McCallum, Barry
**Cc:** Davis, Ben
**Subject:** FW: Plato Contract
**Importance:** High

Debbie approached me today to determine what paperwork was required to exercise the second option year in the Plato contract. When I read the existing contract it clearly stated that PWGSC required a minimum of 15 days to process the extension. As there are less than 15 days remaining in the contract I signed the requisition as A/Director, PSP and faxed it to PWGSC. Below is an e-mail from Debbie which outlines that there may be a problem as we didn't submit the paperwork earlier. I am advising you of this situation as I believe this should have been discussed with you as it is a significant amount of money and it has not yet been actioned by PWGSC so presumably we would be able to call PWGSC and discuss this should it be necessary.

Krystine

-----Original Message-----
**From:** Power, Debbie
**Sent:** Thursday, March 18, 2010 3:03 PM
**To:** Davis, Ben; French, Krystine
**Subject:** Plato Contract
**Importance:** High

Hi Krystine,

 I just got a call from Joanne at Public Works (8667). She advises that there may be a problem with it as it takes about 3 or 4 days to get the approvals up the line.

Debbie

*Debbie Anne Power*

**Fisheries and Oceans Canada**
NAFC, 80 East White Hills Rd.
P.O. Box 5667, St. John's, NL A1C 5X1
Telephone: 709.772.4247
E-Mail: debbie.power@dfo-mpo.gc.ca

2

000006

**Peches et Océans Canada**
NAFC, 80 Rue White Hills Est.
C.P. 5667, St. John's (T-N-L) A1C 5X1
Téléphone: 709.772.4742
Courriel: debbie.power@dfo-mpo.gc.ca

3

# McCallum, Barry

| | |
|---|---|
| **From:** | Davis, Ben |
| **Sent:** | Friday, April 9, 2010 3:55 PM |
| **To:** | McCallum, Barry |
| **Cc:** | Power, Debbie |
| **Subject:** | Fw: VDC |

**Importance:** High

Barry,
This is a very significant milestone that we have achieved. Debbie has really moved the project along and has engagement from users across the board. We should soon discuss some of the next steps we need to achieve.

Have a great weekend!

Ben

---

**From:** Power, Debbie
**To:** Davis, Ben
**Sent:** Fri Apr 09 08:57:23 2010
**Subject:** VDC
Good morning Ben,

As a result of my trip to Halifax and in collaboration with the VDC team in Halifax we now have ERIS3G up and running permanently on the VDC. What this means is that the "eye candy" presented to management at the November 04th meeting is real and working.

I will advise and train a couple of the scientists on the VDC so that we can start analyzing the data that's been entered in ERIS3G and run reports etc. This is a very important milestone and one that will allow end user testing of ERIS3G. It will move the project ahead significantly by providing much needed feedback on the accuracy of ERIS3G.

I intend to train and assign two scientists to do this towards the end of the week of April 19th. The reasons for the week and half delay are:

1. I am at the Cabot Building for G110 training all next week and I want to be present for testing of the VDC before passing it on to the scientists.

2. Jane Moore ▓▓▓▓▓▓▓▓▓▓▓▓▓▓. I need her to change security on the VDC account on the DFO NL production server. I won't be able to get her to do this ▓▓▓▓▓▓▓▓▓▓▓▓ I have put in a request. At BIO anybody is allowed to write data to their Oracle database. We do not want this type of setup. It's important for us to maintain the integrity of the data. Allowing anybody to write anything to the database without the data going through the cleansing process defeats our purpose. Therefore I will ask Jane to tweak the VDC account so this is not possible.

3. When Jane advises that she has changed permissions on the VDC account on our Production Server, I or the Plato guys will log in as user and test to make sure we cannot write to the database. Once testing is complete and we're sure the database is secure I will proceed with end user testing (i.e. have the scientists test ERIS3G using the VDC).

1

Let me know if you would like me to proceed differently or have any other suggestions/ideas you would like to see implemented.

Take care
Debbie


*Debbie Anne Power* 🖐

---

**Fisheries and Oceans Canada**
NAFC, 80 East White Hills Rd.
P.O. Box 5667, St. John's, NL A1C 5X1
Telephone: 709.772.4247
E-Mail: debbie.power@dfo-mpo.gc.ca

---

**Peches et Océans Canada**
NAFC, 80 Rue White Hills Est.
C.P. 5667, St. John's (T-N-L) A1C 5X1
Téléphone: 709.772.4742
Courriel: debbie.power@dfo-mpo.gc.ca

# McCallum, Barry

| | |
|---|---|
| **From:** | Davis, Ben |
| **Sent:** | Tuesday, June 15, 2010 10:33 AM |
| **To:** | McCallum, Barry |
| **Subject:** | Scientific computing |

Good morning,

I am finally getting a handle on what has transpired in our scientific computing/data shop in the last 15 years and what we should do about it. Essentially, we have been in maintenance mode within the Branch with an expectation that IMTS, through a Portfolio Manager, would provide all the programming, network, strategic and tactical support that we might need. The model doesn't work. However, its not working across the country either so we are not alone.

While in maintenance mode, we have been paying IMTS for services that don't meet our needs. We have not been able to interject any lifecycle management practices into our servers because we haven't had anyone to maintain the hardware or manage the software and programs that are accumulating. For example, there are 37,000+ SAS programs, 1600+ Fortran programs and 200,000+ user files residing on Athena. We have finally learned that it is Barry Slaney who needed Athena fired up ONCE a year to do a plot on a dot matrix printer and we could have migrated him off of it years ago if we had been managing the file. Athena will now be shut down in a matter of days with a triplicate backup of its data and a Fortran compiler installed on ERIS.

The data entry and archiving process that is the foundation of our work here has not evolved in the last decade or more. There is poor documentation of processes, and there has been little oversight, control or management until recently. There are efficiencies that would be felt across the Branch if we developed new strategies to manage those processes.

We are at a pivotal point with respect to the data migration project and I have been analyzing next steps with an eye to the currant financial situation in the Branch and strategically, into the future. Plato Consulting could be finished in a couple of months with obvious significant savings on that contract in this fiscal year. However, we will have an Oracle database containing most of our data but we would continue to have a shambles all around it.

With respect to SAS software, our region alone has over 50% of the SAS licenses that the department supports. That is logical given that we are a large assessment based group. We have been working with IMTS and Science Sectors nationally to find a better, less expensive way to support both regional and national SAS needs. We are researching and consulting with SAS to develop the most cost effective way to deliver the service and it will form the national model.

Debbie Power and I have been developing a plan for what is needed to get us to our destination. In a nutshell, we need a CS-2 Programmer to (a) clean up 15 years of neglect, (b) provide programming support to the RES and BI users in the Branch (and we have discovered there is a significant demand for this support) and (c) keep us positioned at the cutting edge of computing into the future (e.g..) support for the CSA Office and IT needs for RAPS. We also need a CS-3 Project Manager to maintain the Oracle database because IMTS won't do it, continue getting the last of our data cleaned up and migrated, manage data entry and archiving, establish best practices for the scientific computing group (and my god we need them!), oversee lifecycle management of our hardware, and provide the management and oversight for information management that the Branch needs.

1

To get around the CS classification in Science Sector issue, Kevin McCarthy has agreed to host them on his org chart if we have the FTE's and salary. I would like to create a CS-2 and CS-3 position on his org chart, put Debbie on assignment in the CS-3 and use her CS-1 salary to partially offset the incremental salary cost. The CS-1 position would remain vacant. The CS-2 position would be a full draw on a salary.

The savings we will realize are in the areas of:
- SAS licensing
- Completion of data migration project
- Timeliness of data entry, editing and archiving will reduce OT in the scientific computing group, and within stock assessment groups because data will be available earlier.
- Reduced hardware costs through elimination of unnecessary servers and rationalization of desktop and laptop computers.

I'd like to talk to you further about what we need before I present it to management committee. We have thought a lot and analyzed more about where we are and where we could be. Many improvements in service and cost effectiveness are possible. We just need to champion them through to completion.

Thanks Barry

Ben

Davis, Ph.D
a/Director, Programs, Services and Planning
Science Branch, Fisheries and Oceans Canada
80 East White Hills Road
P.O. Box 5667, St. John's NL A1C 5X1
(709) 772-0560 // fax/télécopieur (709) 772-4818
ben.davis@dfo-mpo.gc.ca
Direction des Sciences.
C.P. 5667, St. Jean TNL A1C 5X1
Government of Canada/Gouvernement du Canada

## McCallum, Barry

| | |
|---|---|
| **From:** | Davis, Ben |
| **Sent:** | Wednesday, July 14, 2010 8:21 AM |
| **To:** | McCallum, Barry |
| **Subject:** | Fw: Data Migration Project |

We're making good progress here and making up for a lot of lost time.

B.

---

**From:** Power, Debbie
**To:** Mowbray, Fran; Nakashima, Brian; Healey, Brian;         Veitch, Pat; Stansbury, Don; Morgan, Joanne; Healey, Brian
**Cc:** Davis, Ben
**Sent:** Tue Jul 13 21:23:29 2010
**Subject:** Data Migration Project
Hi Team,

The Data Migration Project is moving along swiftly. Ben and I have negotiated with Plato and they have offered two extra developers for a month or so to make up for lost time on the contract. This is very good for us. Currently the team is working on the Import Tools for Pelagics. The FFS Import Tool is completed, tested and working fine. So this message will not apply to those who can use this tool.

While I have these extra developers on hand I would like to get a jump on developing the rest of the import tools required. The import tools mimic the utilities you use now for data collection. Would you be able to send me a sample of each kind of data collection tool you use (i.e. access databases, excel etc)? Even if the data itself is not ready, don't worry about that, I only need a sample to develop the tools.

Also could you send me copies of any and all SQL statements you use to extract your data for reporting and analysis purposes? If I have these I can have views already done for you so you won't have to develop them yourselves. Those who do not use databases to store their data, if you could send me a list of the scripts or programs you run to get your reports and analysis the same applies...I will try to have views created for you to bring back the parameters you need for your reports.

Our database, schema, code tables and data dictionary has been updated on the VDC in Halifax since April. Feel free to access it if you wish. I can set up user accounts for you. Just drop me a line.

Thanks team
Debbie

**He has the right to criticize who has the heart to help.**

---

*Debbie Anne Power* ✌️
**Programmer / Project Manager**
**Data Migration Project**

---

**Fisheries and Oceans Canada**
NAFC, 80 East White Hills Rd.
P.O. Box 5667, St. John's, NL A1C 5X1
Telephone: 709.772.4247
E-Mail: debbie.power@dfo-mpo.gc.ca

1

000012

**Programmeur / Chef de projet**
**Projet des migration de données**

**Peches et Océans Canada**
**NAFC, 80 Rue White Hills Est.**
**C.P. 5667, St. John's (T-N-L) A1C 5X1**
**Téléphone: 709.772.4247**
**Courriel: debbie.power@dfo-mpo.gc.ca**

## McCallum, Barry

| | |
|---|---|
| **From:** | Power, Debbie |
| **Sent:** | Tuesday, July 20, 2010 3:37 PM |
| **To:** | Davis, Ben; Davidson, Fraser |
| **Cc:** | McCallum, Barry |
| **Subject:** | RE: Shellfish data for the Oracle database |

Hi Fraser,

Elaine Hynick has supplied sea cucumber and scallops. Waiting on Lobster, Shrimp, crab and squid.

Thanks
Debbie

| | |
|---|---|
| **From:** | Davis, Ben |
| **Sent:** | Tuesday, July 20, 2010 3:33 PM |
| **To:** | Davidson, Fraser |
| **Cc:** | Power, Debbie; McCallum, Barry |
| **Subject:** | Shellfish data for the Oracle database |

Fraser,

As you know, we met with Don Stansbury and Pat Veitch regarding shellfish data that we want to include in the database. Promises were made that the data would be made available to Debbie however, this has not happened yet. We have secured additional people from Plato to complete the data migration project but we are creating our own bottlenecks with the slow arrival of promised data. Could you please give Don and Pat and nudge and remind them that we are still waiting?

Thanks.

Ben

Ben Davis, Ph.D
a/Director, Programs, Services and Planning
Science Branch, Fisheries and Oceans Canada
80 East White Hills Road
P.O. Box 5667, St. John's NL A1C 5X1
(709) 772-0560 // fax/télécopieur (709) 772-4818
ben.davis@dfo-mpo.gc.ca
Direction des Sciences
C.P. 5667, St. Jean TNL A1C 5X1
Government of Canada/Gouvernement du Canada

# McCallum, Barry

| | |
|---|---|
| **From:** | Power, Debbie |
| **Sent:** | Monday, July 26, 2010 10:10 AM |
| **To:** | Davis, Ben; Davidson, Fraser |
| **Cc:** | McCallum, Barry |
| **Subject:** | Update on shellfish |

Good day,

  This is to give you an update on the Shellfish data that the Data Migration Project has been trying to acquire.  We were looking for:

1. Lobster
2. Shrimp
3. Crab
4. Squid

  I met Pat Veitch in the cafeteria this morning and we discussed delivery of these.  This is very important now as I have extra help from Plato for one month to develop these import tools.

Pat advises that he will have Shrimp and Lobster for me by noon today.

Crab is in db3 format and he's the only one who knows how to program in db3.  Currently they are planning on moving this to MS Access but it looks like it will be in db3 for the rest of the year.  We will have to revisit this.  It doesn't make sense to put developer time into a program they are not planning to continue with and they do not have their go forward database developed yet.

They are not currently collect Shrimp so an import utility can't be developed for this data structure.  We will however be able to load historical Research Squid data into the new Oracle database manually without an import tool.

So Fraser, no follow up is required on your part unless I do not get Shrimp and Lobster today.

Thanks all - have a great week
Debbie


**"He has the right to criticize who has the heart to help."**

*Debbie Anne Power* 🕊

**Programmer / Project Manager**
**Data Migration Project**

**Fisheries and Oceans Canada**
**NAFC, 80 East White Hills Rd.**
**P.O. Box 5667, St. John's, NL A1C 5X1**
**Telephone:  709.772.4247**
**E-Mail:  debbie.power@dfo-mpo.gc.ca**

1

000015

**Programmeur / Chef de projet**
**Projet des migration de données**

**Peches et Océans Canada**
NAFC, 80 Rue White Hills Est.
C.P. 5667, St. John's (T-N-L) A1C 5X1
Téléphone: 709.772.4247
Courriel: debbie.power@dfo-mpo.gc.ca

# McCallum, Barry

| | |
|---|---|
| **From:** | Tillman, Joe |
| **Sent:** | Wednesday, September 15, 2010 10:15 AM |
| **To:** | McCallum, Barry |
| **Subject:** | Data Migration |

This topic will come up on Friday and I thought it would be best that I give my input directly to you. I have tried in the past several months to get MC's attention on what appeared to be a shift in our strategy. As you may recall, this strategy received overwhelming support from MC and SHs last November. In a nutshell we had agreed to migrate the MS data for gfish, shellfish and pelagics, build the FFS replacement, build the back-end tools (VDC, R and other access tools) and then test the 'Dickens' out of it before we cut over to production.
I'll freely admit that I may have an attachment to the initiative, although I rationalize that by recalling the many hours of 'blood, sweat and tears' that I devoted to what was once considered a top priority for the branch. I am also sensitive to our fiscal issues but you'll recall that I did question our ability to cancel the contract in mid-year but I now hear that it's a deferral and I would find it difficult to believe that we have successfully achieved the milestones that we set for the project. The question I would ponder (but not ask at MC) is " What elements of this strategy have changed given the 'deferral' of the Plato contract and have we given sufficient thought and discussion to the consequences of the decision?"

Again, I'm not interested in a debate (or another rebuff) and that's why I'd prefer to offer these comments to you alone in this manner. Just as info I've attached the 2 decks we've used in the past to communicate the Branch Strategy. I can add a sub-text in person if you wish to talk about it further.

**DFO Presentation**    **Data Mgmt for**
**final.pdf**        **NSDC.ppt**

J.M. Tillman
Advisor to the Regional Director
Science Branch
NL Region
Ph.: (709) 772-4355
Fax: (709) 772-6100
e-mail: *Joe.Tillman@dfo-mpo.gc.ca*

1

# Data Migration Project

Presentation to Management Committee

November 4th, 2009

Plato
PARTNERS IN IT SOLUTIONS

# Data Migration Project

**Purpose:**

- Migrate the NAFC Science data from flat file archives to a modern database

**Benefits:**

- Maximize the usefulness of existing data through standards
- Improve public access and interconnectivity of data assets
- Make data-dependent processes simpler and more efficient
- Enable the use of modern tools that provide easier data access and allow deeper analysis

Plato
PARTNERS IN IT SOLUTIONS

# Background

- **Phase I: Strategy & Planning (January to March 2006)**

  – Examined assets and processes to identify key issues

  – Developed a multi-year strategy for the project

- **Phase II: Start of Migration (November 2006 to March 2008)**

  – Refined scope of included data

  – Analyzed and addressed data format issues

  – Built the migration process

  – Designed initial database schema

**Plato**
PARTNERS IN IT SOLUTIONS

# Background cont'd

- **Phase III: Migration (June 2008 to Present)**

  – Analyzed and addressed data relationship issues

  – Completed enhancements to migration process and database schema design

  – Developed editing, analysis, and reporting tools to support existing business processes with new system

Plato
PARTNERS IN IT SOLUTIONS

# Basic Data Model

**Series** – Multi-year data set for long-term analysis (e.g. Fall Survey)

**Batch** – Samples collected together (e.g. Trip by research vessel)

**Sample** – Organisms collected together (e.g. All fish from a trawl)

**Catch** – Sample for one species (e.g. Cod from a trawl)

**Organism** – Data for individual organisms (e.g. Fish length)

# Data Analysis & Cleansing

- Analyzed records for conditions that would prevent migration or might indicate other problems

- Analyzed relationships between records in search of problematic conditions

- Worked to resolve as many issues as possible
  - Cross-referenced data to original sources both electronic and paper
  - Gather missing codes to fill gaps in code tables

Plato
PARTNERS IN IT SOLUTIONS

# Database Schema

- Worked with scientists and other stakeholders on data subcommittee to design new schema

- Drew information from BIO and BIOCHEM database schemas, as well as examining the FFS data collection schema

- Used an organism-centered data model to provide better support for novel analysis in future and to remove uncorrelated data 'silos'

Plato
PARTNERS IN IT SOLUTIONS

# Meta-Data

- Gives context and reduces misinterpretation

- Had typically been stored outside the archive and was often outdated or missing

- New DB has extensive meta-data support

  – Can be stored in the archive

  – Once entered, it will not be lost or separated

  – Electronic documents can be stored with data

  – Freeform text notes can be attached to any record

Plato
PARTNERS IN IT SOLUTIONS

# Migration Process

- Four stages with a total of 92 processing steps

  — **Stage 1**: Initial data load (all data as text)

  — **Stage 2**: Validate types and codes; apply corrections; load to correctly typed tables

  — **Stage 3**: Validate keys and relationships; load to partially normalized, related tables

  — **Stage 4**: Build target database structure; transform and load data to fully normalized target schema; unify data to remove remaining duplication

Plato
PARTNERS IN IT SOLUTIONS

# Data Migration by the Numbers

| | |
|---|---|
| **3,151,093** | **Source records loaded in Stage 1** |
| 264,331 | Duplicate records identified and removed<br>8.4% of source records |
| 79,138 | Non-duplicate records that could not be loaded<br>2.5% of source records; 2.7% of non-duplicate source records |
| 1,538,158 | Records adjusted for issues found during analysis<br>48.8% of source records |
| 3,307,927 | Individual fields adjusted for issues found during analysis<br>Average of 1.05 fields per source record |
| **2,807,624** | **Records migrated to new archive in Stage 4**<br>**97.3% of non-duplicate source records** |
| 96,126,705 | Target records in new archive after migration |

Plato
PARTNERS IN IT SOLUTIONS

# Analysis & Reporting

- Previous Archive
  - Data access was laborious and error-prone
  - Data had to be marshaled from dozens of files
  - Data sets had to be preprocessed to handle anomalies

- New Archive
  - All data and meta-data in one place
  - Anomalies have been corrected
  - Access with existing tools is simpler & more efficient
  - New tools are enabled

Plato
PARTNERS IN IT SOLUTIONS

# New Tools – STRAP in R

- R is a Statistical Computing Environment

- Continues work begun with the previous archive

- STRAP1 and STRAP2 analysis have been implemented in R

Plato
PARTNERS IN IT SOLUTIONS

# New Tools - VDC

- The Virtual Data Centre is a web-based data access and analysis tool developed at BIO

- In cooperation with BIO, access to the new archive has been established through the VDC

- It provides full database query functionality, as well as data visualization capabilities

- It will provide a conduit for data in the new archive to national and international systems

Plato
PARTNERS IN IT SOLUTIONS

# Future Directions

- Continue work to adapt and develop analysis and reporting tools

- Conduct comprehensive system testing

- Migrate more data sets, including Pelagic and Invertebrate data

- Support business process changes, such as meta-data collection and code management

Plato
PARTNERS IN IT SOLUTIONS

Fisheries and Oceans Canada    Pêches et Océans Canada

# Data Management Priorities in the NL Region: Data Migration

Canada

Fisheries and Oceans    Pêches et Océans
Canada    Canada

# NL Data Priority: Migration of 40 Years of Data from Flat Files to A Relational Database

- Phase I – identified key issues and developed a high-level plan

- Phase II – analyzed and addressed data format issues, designed target database, developed migration process, and performed initial migration

- Phase III – analyzed and addressed data relationship issues, refined database schema and migration process, developing analysis and reporting tools

Canada

2

Fisheries and Oceans   Pêches et Océans
Canada                 Canada

# NL Data Priority: Migration of 40 Years of Data from Flat Files to A Relational Database

Major Accomplishments/Activities:

- Identified key issues with the current system

- Developed a multi-year strategy for enhancing the Multi-Species Trawl Survey data set

- Refined scope of included data

- Analyzed and addressed data format issues

- Built the migration process

- Designed database schema

- Created FFS import tool

- Analyze and address data relationship issues

- Enhance migration process, database schema design, and FFS import tool

- Develop editing, analysis, and reporting tools to support existing business process with new system

Canada

3

Fisheries and Oceans    Pêches et Océans
Canada                  Canada

# The Scope:

| | |
|---|---|
| **3,151,093** | **Source records loaded in Stage 1** |
| 264,331 | Duplicate records identified and removed *8.4% of source records* |
| 79,138 | Non-duplicate records that could not be loaded *2.5% of source records, 2.7% of non-duplicate source records* |
| 1,538,158 | Records adjusted for issues found during analysis *48.8% of source records* |
| 3,307,927 | Individual fields adjusted for issues found during analysis *average of 1.05 fields per source record* |
| **2,807,624** | **Records migrated to new archive in Stage 4** *97.3% of non-duplicate source records* |
| 96,126,705 | Target records in new archive after migration |

Canada

Fisheries and Oceans     Pêches et Océans
Canada                   Canada

# The Future

- Workload efficiencies: Reduction in many hours of data preparation, analyses and data requests

- New data acquisition tools and applications (zonal initiative)

- Accessibility: Linkages to VDC (testing in progress) and OBIS

- Safeguarding valuable asset

Canada

5

# McCallum, Barry

| | |
|---|---|
| **From:** | Davis, Ben |
| **Sent:** | Tuesday, September 21, 2010 11:04 AM |
| **To:** | McCallum, Barry |
| **Subject:** | FW: Plato contract |

FYI

---

**From:** Power, Debbie
**Sent:** Tuesday, September 21, 2010 10:32 AM
**To:** Davis, Ben; French, Krystine
**Subject:** Plato contract

Hi Ben/Krystine,

Plato has agreed to the contract amendment. They would like to know how long it will take us to draft up the papers. I will begin the 9200 form today.
Update on the Data Migration Project itself:

1. ░░░ and ░░░will return to Plato tomorrow.
2. ░░░went back to Plato for a couple of days for training this week. He will remain there.
3. ░░░will stay on until October 08th. ░░░░░░░░░░░░░░has been working with him on the Crab collection tool and the Crab import tool. This will give us time to complete this development segment.

Could you let me know if you require anything else from me please?

Debbie

**"He has the right to criticize who has the heart to help."**

---

*Debbie Anne Power* ⁀ઇ
**Programmer / Project Manager**
**Data Migration Project**

---

**Fisheries and Oceans Canada**
**NAFC, 80 East White Hills Rd.**
**P.O. Box 5667, St. John's, NL A1C 5X1**
**Telephone: 709.772.4247**
**E-Mail:** debbie.power@dfo-mpo.gc.ca

**Programmeur / Chef de projet**
**Projet des migration de données**

---

**Peches et Océans Canada**
**NAFC, 80 Rue White Hills Est.**
**C.P. 5667, St. John's (T-N-L) A1C 5X1**
**Téléphone: 709.772.4247**
**Courriel:** debbie.power@dfo-mpo.gc.ca

## McCallum, Barry

| | |
|---|---|
| **Subject:** | Data Migration Project |
| **Location:** | Barry McCallum's Office |
| | |
| **Start:** | Tue 9/20/2011 10:00 AM |
| **End:** | Tue 9/20/2011 12:00 PM |
| **Show Time As:** | Tentative |
| | |
| **Recurrence:** | (none) |
| | |
| **Meeting Status:** | Not yet responded |
| | |
| **Organizer:** | French, Krystine |
| **Required Attendees:** | McCallum, Barry; Mansour, Atef A H; Manning, Joe; Davis, Ben; Tillman, Joe; Senciall, Dave; Power, Debbie; Carroll, Marlene; Brodie, Bill |

When: September 20, 2011 10:00 AM-12:00 PM (GMT-03:30) Newfoundland.
Where: Barry McCallum's Office

Note: The GMT offset above does not reflect daylight saving time adjustments.

*~*~*~*~*~*~*~*~*~*

Further to the September 15, 2011 Management Committee Meeting it was agreed that we would meet early next week to discuss the Data Migration Project. Debbie will provide an update as to the status of the project. We now require a decision regarding which direction to follow for the remainder of the Fiscal Year. Attached is the Business Case for discussion at the meeting.

Business Case for Logical_States_2... Physical_States_...
Data Migrati...

1

# MINUTES

## Science Management Committee Meeting
### Sept 15th, 2011
### 9:30 am – 12:30 pm

**In Attendance:**

| | |
|---|---|
| Barry McCallum (Chair) | Ben Davis |
| Atef Mansour | Bill Brodie |
| Joe Tillman | Joe Manning |
| Marlene Carroll (Minutes) | Krystine French |

1. **Call to Order**
   The meeting was called to order at 9:35 am.

2. **Review of Agenda**
   Agenda was reviewed and accepted. Items added included Vessel schedule, Firearms and Accommodations committee.

3. **Staffing (Krystine French)**
   Management Committee had a discussion on divesting determinate staffing to the Division level with monthly reporting to Management Committee.

4. **Integrated Financial and Operational Planning (Joe Tillman)**
   Discussion was held on what this new Planning tool will look like and what is expected of Division Managers. Krystine will determine what reports Division Managers will need to run and organize a training session for Division Managers.

5. **Performance Reviews and Learning Plans for 2011-2012 (RDS)**
   RDS asked Management Committee members to try and get these completed. Extraction date is Friday, September 23rd.

6. **Training Requests (Atef Mansour)**
   Discussion was held at the BUMCC meeting re: requirement for Health and safety training for Science staff. OSH committee will deliver the training. Atef will follow up.

7. **Credit Card/contracting Issues (Atef Mansour)**
   There are some issues with staff using credit cards for purchases when there should be a contract in place before purchase is made. Atef feels that training is required for staff. Kystine will speak with Contracting and Procurement staff and set up a training session.

8. **NL EBSA Steering Committee meeting (Bill Brodie)**
   Steering Committee meeting was held on September 9th with Bill as chair. The terms of reference and objectives for the committee and the CSA process were reviewed.

Extensive discussions were held on process, methods, and timelines. The committee came up with key conclusions for the process.

## 9. Plato contract (Debbie Power)
Deferred until next week.

## 10. Vessel Schedule (Krytine French)
Krystine distributed a complete list of vessel requests for the 2012 survey schedule. RDS asked Krystine to discuss with George and ask that he analyze with comparison to last years schedule.

## 11. Firearms (Krystine French)
There is some Science staff that needs to be identified who requires the use of firearms in their job description. Firearms training is required for all staff concerned and this will be captured on their work description as "Skills required".

## 12. Accommodations Committee (Krystine French)
Krystine gave an update on meeting she attended last week. The move is still on-going for Science staff.

## 13. Roundtable
**Bill Brodie** – Attending NAFO meeting next week.

**Joe Manning** – Reported that Matthew trip is complete and staff is back to the office.

**Krystine French** – Antoine Sioufi (Manager, Intellectual Property & Partnering) will be visiting the Region next week and will be attending Management committee meeting. He will also meet with all staff in the Auditorium.

**Atef Mansour** – A cable broke on the CTD on the Needler. However, the CTD may be retrievable.

## 14. Meeting adjourned at 12:35 pm.

000040

<u>**Business Case for Management Committee**</u>

<u>**DATA MIGRATION PROJECT**</u>

<u>**Background:**</u>

The Data Migration Project is a regional initiative to take all the Science data currently
residing in flat files and excel spreadsheets on ERIS and move them into an Oracle
database. This supports the National initiative to make all Science data discoverable.

The project began in 2006 and has continued through to present. The contractor for the
project is Plato Consulting.

<u>**Analysis / DFO Comment:**</u>

The overall goal of this contract was to clean Groundfish, Pelagic's and Invert (Shrimp
and Crab) data and enter them into a new Oracle database (ERIS3G) and finally to create
import tools for this data to allow DFO staff to clean and import future data into ERIS3G.

During the 8 months from February to September 2010 the following deliverables were
completed, tested and debugged.

Cleaning of Groundfish data was completed. This dataset is 95% clean.
Cleaning of Pelagic's data was completed.
Cleaning of Invert's data was completed.

1. FFS Import Utility: This utility had been developed earlier in the project. It was
   discovered in February that this tool was not functional. The tool was revised,
   tested and debugged. Groundfish data was successfully loaded to the new Oracle
   database on September 22 using the FFS Import Utility.
2. Capelin Import Utility: This utility is completed.
3. Sand Lance Import Utility: This utility is completed.
4. Shrimp Import Utility: This utility is completed.
5. Crab Import Utility: This utility is completed.

<u>**Work Remaining:**</u> The following import tools were not completed during the 8 month
period either because time ran out or the datasets weren't ready on time.

1. Herring (data wasn't ready in time but is ready now).
2. Brian Nakashima's data (data wasn't ready, mimics Capelin so it will not be hard
   to develop an import tool.)
3. Scallops (time ran out). Note this dataset was not part of the deliverables on the
   contract.
4. Squid (time ran out) Note this dataset was not part of the deliverables on the
   contract.

5. Sea Cucumber (time ran out) Note this dataset was not part of the deliverables on the contract.

## Recommendations / Next Steps:

DFO Science still has approximately 190K remaining on the Plato Consulting contract. I would like to recommend a different approach to this database development.

It is unusual to develop a database structure without an interface for users to interact with the database that allows them to enter data (choose an import tool and import new data) and retrieve information on the data (print reports, analysis, plots etc).

At this point in the project we have two alternatives:

## Option 1:  Continue to complete the import tools

We can continue as we have been and finish creating import tools for the datasets listed in the "Work Remaining" section above.

## Option II:  Move to different approach

Use the remaining money on the Plato contract to layout a plan for the new Oracle database.

- Design document for a user interface that will allow the users to actually use the database
- Design document for reports that the user can generate automatically (NAFO reports, RAP reports etc). The parameters for these reports are constant so this can be programmed to produce a report and an excel spreadsheet that users can run SAS programs on to analyze the data.
- Design document for an utility that will allow the users to plot their data

The deliverable from the above will be very thorough Project Plan that will detail what needs to be developed (deliverables), how it will be developed, timelines, costs, resources etc. This document will be precise and include stakeholders, steering committee recommendations, work groups etc.

This will give DFO to a complete and precise "game plan" so that this work can be contracted out when money is available or done in-house if there are budget constraints. Each deliverable will be clearly defined so that the developer (whether in-house or contractor) will have a clear path laid out for the successful completion of the Data Migration Project.

## Option III: The same as Option II above but with Science Branch creating the Project Plan document and setting the way forward

This would mean that Science Branch would do the work outlined in Option II above and "buy out" the existing Plato contract. Science Branch has the in-house knowledge to produce a detailed Project Plan for the way forward.

Of the above options I would like to recommend Option III.

**Advantages for Science Branch:**

With Option III, once the interface is completed, the end users can begin to use the database and new datasets can also be migrated into the database. This has many advantages.

1. We have many import tools developed and data entered in the new Oracle database. If something has been overlooked or there is a design flaw it won't be picked up until the end users actually begin interacting with the database.
2. The further into development we get before we have user testing the more costly and time consuming it will be to correct flaws in the database.
3. The sooner the users can interact with the database the sooner we can get their feedback. This will allow us to fix bugs early and not carry errors throughout the development of other datasets.
4. We can catch anything important that may have been overlooked or not thought of while it is easier to correct or add. Especially, if it results in a re-construct of the database.

**Attachments:** Attached you will see two dataflow diagrams. The current state in these documents reflects Option I. The future state reflects Option II and Option III.

## Current State

**Data Collected from:**
Capelin
Capelin Spawning
Crab
Groundfish A
Groundfish Fecundity
Herring
Sandlance
Scallop
Sea Cucumber
Shrimp
Starfish
Unknown

Data Recorded on Handwritten Data Sheets

Data Manually entered

VDE Image Plus

Edit Reports created by Data Archivist using SAS Programs

Reports reviewed by Data Custodian

Data Corrections Needed

Reports are accurate

Flat Files archived on ERIS (OpenVMS) by Data Archivist

Data Entered into FoxPro database via FFS Utility while on Vessel

Once on land, Flat files sent via Email

Edit Reports created by Data Archivist using SAS Programs

Reports reviewed by Data Custodian

Data Corrections Needed

Reports are accurate

Flat Files archived on ERIS (OpenVMS) by Data Archivist

## Future State

**Web Based Application**
- Secure
- Per User
- One System – All Species
- Instant Validation of Data
- Accessible Anywhere (Internet and Synchronization System)
- Links to other systems

Data Entry – All Species
Data Retrieval
Data Edits
Task Lists
Workflow (Data Signoff)
Generic Reporting

**Business Process Engine**
- Secure
- Data Validation Rules Engine (Replicate SAS Programs plus additional)
- Customizable
- Trained staff only

Data Input Administration
Data Input Rules
Species – Add or Removal
User Account Administration

**Database**
Imported/Migrated Data resides in secure Oracle database

Managerial/Statistical Report Generation
Data Extraction
System Integration

## Current State

**Database**
Migrated Data
Resides in a Secure
Oracle Database

ERIS
(OpenVMS)

Handwritten Data Sheets
FFS Utility (FoxPro Database)
Data Migration Import Tools

**Data Sources**
Capelin                          Sandlance
Capelin Spawning          Scallop
Crab                               Sea Cucumber
Groundfish A                  Shrimp
Groundfish Fecundity     Starfish
Herring                          Unknown

## Future State

Copy of
Production Data

Business Logic,
Reporting and Data
Extraction Security

Reporting and
Data Extraction

Canned Reports, Ad Hoc Reporting,
and Report Designer

Secure
Oracle
Database

Business Rules to guide/enforce
Data Entry and Secure Retrieval,
User Accounts, System
Integration

Web Application for Data
Entry, Validation and
Retrieval

Data
Synchronization
Server

User connected
to GNL via VPN or
LAN

Users connected to
internet via LAN, Wireless,
Satellite etc.

Internet

User returns to Office
or Home.
(Obtains Internet
Connection)

**Usage from the Field**

Users performing data entry
without internet access using
a local application on Laptop
or PC running all Data
Validation rules